Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Computer Science 13 (2012) 53 – 59

Procedia Computer Science

Proceedings of the International Neural Network Society Winter Conference (INNS-WC 2012)

A Framework of Multi-Stage Classifier for Identifying Criminal Law Sentences

Sotarat Thammaboosadee[a,*], Bunthit Watanapa[a], Nipon Charoenkitkarn[a]

[a]*Data and Knowledge Engineering Laboratory (D-Lab), School of Information Technology, King Mongkut's University of Technology Thonburi, 126 Pracha Utit Rd., Bangmod, Thungkru, Bangkok, 10140, Thailand*

Abstract

This paper proposes a framework to identify the relevant law articles consisting of sentences and range of punishments, given facts discovered in the criminal case of interest. The model is formulated as a two-stage classifier according to the concept of machine learning. The first stage is to determine a set of case diagnostic issues, using a modular Artificial Neural Network (mANN), and the second stage is to determine the relevant legal elements which lead to legal charges identification, using SVM-equipped C4.5. The integrated multi-stage model aims at achieving high accuracy of classification while reserving "arguability". Hypothetically, mANN handles well for digesting complexity in case-level issues analysis with acceptable explanatory power and C4.5 addresses the lesser extent of contingency and provides human-interpretable logic concerning the high-level context of legal codes.

Keywords: criminal law, data mining, neural network, decision tree, legal reasoning

1. Introduction

The literature of machine learning in the legal domain mostly focuses on the common law system [1, 2, 3]. Developing a Decision Support System (DSS) for the domain of civil law, therefore, will be beneficial to those who are involved or interested in the legal system in many countries, including Thailand. The DSS of our interests in the civil law system shall provide knowledge in terms of validated case attributes and their relations

*Corresponding author. Tel.: +66-8140-23627.
E-mail address: sotarat@sit.kmutt.ac.th

1877-0509 © 2012 Published by Elsevier B.V. Selection and/or peer-review under responsibility of Program Committee of INNS-WC 2012
Open access under CC BY-NC-ND license. doi:10.1016/j.procs.2012.09.113

to the possible sentences which can provide an inductively logical basis for learning and understanding this legal system.

Hypothetically, in the judicial process of civil law system, the facts in criminal case would be collected, investigated and induced into a set of clue that suggests the related elements in law articles. To reflect this concept, this paper proposes a two-stage data analysis: (i) criminal facts are collected and formalized into fact-level data, then, classified into abstraction of case-level data, alternately called the case diagnostic issues and, (ii) the case-level data are analyzed into legal elements, or legal-level data, of the correspondingly applicable articles for identifying a range of possible sentences, ranges of a punishment and also their exacerbate or mitigation due to the additional provisions.

Generally, the process of legal reasoning or decision making in judicial process, specifically to the court level, requires the induction of the result for the advantages of cases explanation and argumentation. For case level identification, the modular ANN is designed so that its architecture imitates the process of low level analysis. In legal elements identification, the adapted decision tree C4.5 was selected as the model builder due to its inductive ability, which is a necessity for such a high level reasoning where human interpretability is required.

The remainder of the paper is organized as follows: the next section shows relevant past works in the law domain on classification, especially to the modular ANN and rule-based method such as decision tree. Section three introduces the methodology in the classifying process which includes the overall architectural design. Finally, we conclude the paper and suggest the directions in applying of the framework as future research.

2. Related works

There are several approaches to develop a human-interpretable legal intelligent system. A decision tree is one of the simplest but yet provides powerfully interpretability. However, the decision tree normally does not provide the best classification accuracy but it could be used to analyze insights into the interaction of variables [4]. Gutierrez et al. [5] proposed a crime report prediction system driven by decision tree as data mining method. This system purpose is to predict unreported crime by C4.5 decision tree algorithm. However, the accuracy of the model was quite low (around 56-75%) but their advantage is the elimination in thousands of attributes and construction of an explainable model.

To compare different data mining methods, Yang et al. [6] constructed four independent classification models: random forest, decision tree, artificial neural network (ANN) and support vector machine (SVM) in a prediction system of an offender affiliation and murder victim. Their experiments showed that the random forest method had the lowest error rate while SVM and ANN had the best accuracy in the training set. However, these resulted in the worse accuracy in testing set because of overfitting issues. Interestingly, the decision trees generated the highest error rates but they provided the benefit in the sequential factor analysis and offered knowledge insight.

Recently, Stranieri and Zeleznikow [7] proposed the Split-Up system that determines the final judgment for Australia's family cases, specific to the percentage of assets splitting for divorces. Various artificial intelligence and knowledge discovery techniques were applied, especially, to the reasoning system. For example, a neural network was applied to discover the value of the heuristically defined legal factors in their factors hierarchy, which was used to imitate the factors that the courts usually have to sequentially determine. Expert-constructed fuzzy rules were also used to consider the vague legal terms, such as the level of assets or a parent's maturity. Finally, the complete system was put online and interacted with users with the help of a decision tree [8]. However, this work is different from the domain of our framework. The required final results in civil law domain, especially in criminal law, should specify all identified legal elements based on the legal code and map onto charges.

Another research that applied ANN to legal tasks was proposed by Oatley, Ewart and Zeleznikow [9]. This

evidence-based forensic science application used data from various judicial sources to discover the criminal factors in classifying types of crime. Simple feed-forward back-propagation ANN architecture was applied. However, the lack of inferable structure such as modularity or tree limited the effectiveness of this application to a detection system. It demonstrated that ANN is an effective tool with acceptable accuracy range of 75-92%, for determining legal factors.

Theresa and Raj [10] also reported an effective usage of ANN in legal domain. They applied a back-propagation ANN to classify murder cases. The input data were processed and summarized by experts. The three output classes were separated by kinds of punishment. Although the result is satisfactory due to the comparison with expert's decision, the lack of inference and overly abstract outputs are areas of prospective improvement.

In other domains, the modular ANN has also been proved as an effective tool for classification. Watanapa et al. [11] proposed a sieving ANN to classify emotion-based movie clips into three groups: excite, joy and fear. Their ANN architecture consists of two sequential modular sub-networks. The accuracy of this modular ANN is greater than the result of traditional ANN.

In the next section, we will discuss the methodology for this system and its limitation in terms of scope of application and data availability.

3. Methodology and scope of application

In this paper, a frame work of two-stage classifier is proposed for the benefits of specifying prospective legal elements in criminal law (specifically for the life and body section), based only on the collection of the past cases in a specific sovereign state, e.g. Thailand. Normally, an incident in criminal law case contains a large number of fact-level attributes which are used to determine the legal charges. To impose a learning machine inferable of a systematic judicial system, these factors need to be conceptualized into an intermediate level of abstraction, called case diagnostic issues (case-level attributes). A court usually relies on the case diagnostic issues to cautiously measure the legal elements (legal-level attributes) that can be mapped onto the legal charges as strictly defined in the codes and the statutes.

In respect of such a hypotheses requirement, a multi-layer classifier is proposed to determine the legal charges and punishment ranges. The first layer would identify the case-level attributes based on obtained fact-level attributes. This is very important and requires computing schemes that can comprehend the subtle way of human judgment. This classifier level was built with Artificial Neural Network (ANN) algorithm equipped with modular structure. By empirical experiment, we found that even the traditional ANN is more promising when compared with tradition approaches of decision tree and naïve Bayes. The accuracy of ANN results is in the range of 65-82%, while the accuracy of those traditional approaches is in the range of 58-68%.

The second layer is the legal-level attributes identification based on the case-level attributes acquired in the first stage. The C4.5 decision tree method was selected because of its inferable result. Finally, the identified legal elements are then mapped to the legal charge codes which provide the sentences and ranges of punishments. Fig. 1 shows the architectural design of our proposed framework.

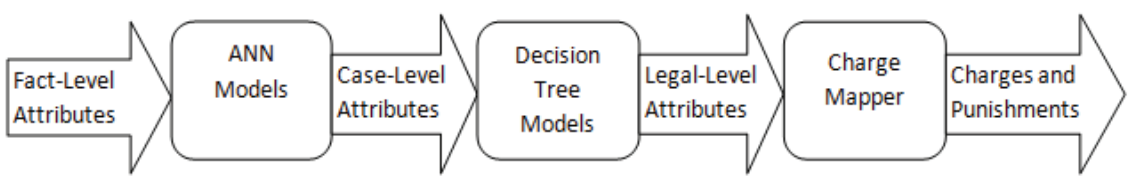

Fig. 1. The architectural design of the proposed system.

Anyway, not all of the fact-level attributes were fed into the case-level classifier because some attributes can be directly mapped using explicit rules, according to empirical experiments. In the same manner, only a certain number of case-level attributes are required for decision tree analysis. In our experimental system, the total 99 fact-level attributes are used to identify 25 case-level attributes. Subsequently, those obtained case-level attributes would be fed to legal-level attributes identifiers, which is deployed as decision tree models in this framework.

The experiment for this study is planned to be performed using the offences against life and body part of the criminal law in Thailand. Unluckily, we have obtained only a limited number of 150 cases only. Considering the voluminous attributes required as input for the classifier in this system, an effective factor analysis method is hence required [12]. In our framework, two schemes of Principal Component Analysis (PCA) [13] and Support Vector Machine (SVM) attributes ranker [14] are proposed to reduce the input dimension without loss of determining power of the classifying system. To further squeeze the performance, the test is planned to be executed under the 10-fold cross validation method and data will be cleansed for ensuring standardization and integrity [15].

In the next subsections, we give an overview of the design of the models in both classifiers, and end up with an overview in experiment and evaluation.

3.1 Case Diagnostic Issues Identification

The selected classifier in this stage is the Modular Artificial Neural Network (mANN) with feed-forward and back-propagation architecture [16]. In our mANN model, prospective extracted inputs are 99 raw attributes. Considering the target classes, the case-level attributes have been categorized into 25 groups. The mANN emulates the divide-and-conquer method in a way that the final outcome of the system is consolidated from the other ANN modules [17]. To express the relationship between the facts and the diagnostic issues in views of mANN, a modular sample of Act_Force_Level case-level attribute (C1: the level of force in the action) is shown in Fig 2.

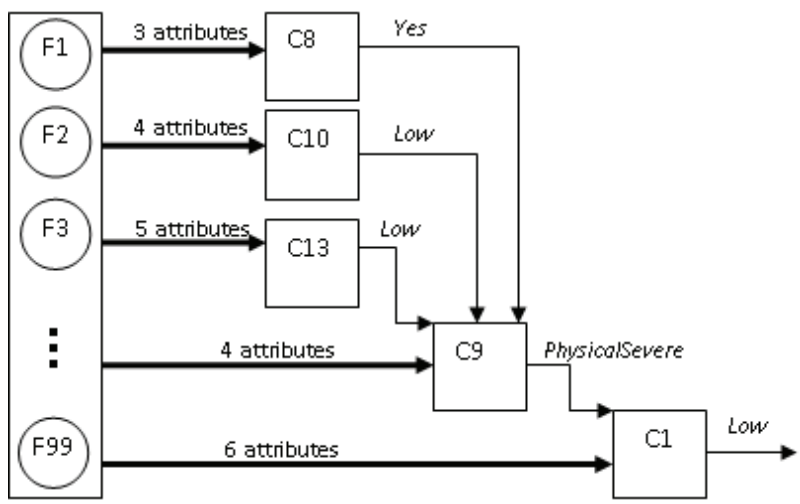

Fig. 2. The modular structure example of the case diagnostic issues identification for Act_Force_Level.

The C1 consists of a modular C9 (severity of the result), which also contains 3 sub modules: C8 (important of organs), C10 (severity of sore) and C13 (severity of weapon). The modular structure in this sample is

designed based on the procedural legal investigation logic. Such the hierarchical relationship could be validated by statistical correlation based on the training data. With non-modular structure, the determination of C1 will face high dimension of 22 variables and would get stuck in the problem of over-fitting due to the limited training data set.

3.2 Legal Elements Identification

This classifier is to determine the legal level attributes based on the case-level attributes achieved from the mANN in the first stage. The possible sentences will be forecasted, hence, the interpretation of why specific articles are chosen have to be inductively arguable based on the specified set of case diagnostic issues. In this level, the inductive inference engine is needed to emulate the reasoning process. In our system, the decision tree with C4.5 algorithm [18] is selected as the classifier. The C4.5, an extension of the ID3 algorithm [19], is an algorithm used to generate a decision tree which is constructed upon the concept of information entropy. At first, the algorithm finds the splitting with the highest normalized information gain on each attributes. Then it creates a decision node with the selected node and splitting. The algorithm recurs at the obtained sub-list by splitting on the selected attribute and adding those nodes as child nodes. The improvements of the C4.5 from ID3 are missing values replacement, both continuous and discrete attributes handling and tree pruning.

The legal level attributes are defined based on the classes derived from analyzed criminal law theory constructed based on criminal law theory and elements in law articles [20]. Since the scoped criminal law articles in our work are only life and body section, the values of the structuralized legal level attributes, which are the target classes on this classifier, were also scoped to the law articles appeared on this section.

For an idea of the decision tree in this legal reasoning system, let's consider a hypothetical induction system of Awareness class, which refers to the awareness level of the offender, as shown in Fig 3. Each tree node of this probable decision tree is labeled in decoded description. The decoding is transformed from the encoded version which was predefined in the numeric values. This decision tree could be constructed by C4.5 with SVM attributes ranker preprocessor as mentioned earlier. To demonstrate the usage of this decoded inference model, we give an example of determining anger mindset in an offender. According to Fig 3, the initial condition is that an action is necessary or not. If the action has no necessary reason, an offender will be judged as fully aware. On the other hand, if the action does not have sufficient reason (low necessity), is performed with the discontinuous time period and without planning, the offender will be judged as partially aware, which will mitigate the sentence. Otherwise, the offender will be judged as fully aware which cause the offender must be liable and punished with the original specified sentence in law article [20].

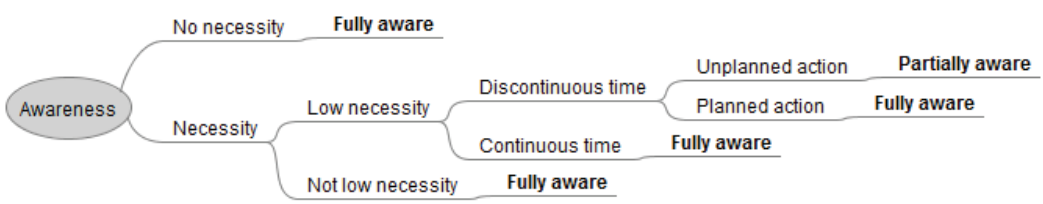

Fig. 3. A sample of decision tree for "Awareness" class.

4. Discussion

As noticeable from the illustration above, the two-stage process gains advantage in this model since the first stage of investigation process that transforms case facts into abstracts or instruments (for further referring legal

elements/articles) is complicated and relies on human judgments. The ANN is a reliable method for imitating the logic in human judgment. Actually, other machine learning techniques could replace ANN. For example, SVM or Bayes network can be good candidates, provided there is large training data set.

The second stage of identifying legal article is different story. The key issue of this stage is the inferring ability (for example, why a particular legal element is chosen for the case abstracts). It is so systematic that the decision tree can reliably handle it and its induction process will automatically give a reason for legal identification with a tracing path of inference. The induction core of decision tree yields a rule-based feature that helps to interact with human users with a more interpretable reasoning.

5. Conclusion

A framework of multi-stage classifier is proposed to identify sentences and the defined range of punishment according to the legal domain of interest which is the criminal law in the civil law system which the sentences are relied on law articles. To support the framework, the data layers are separated into three levels: fact-level (facts in cases), case-level (case diagnostic issues) and legal level (legal elements). Based on the acquired fact-level data, the first classification stage identifies the case-level attributes, which may be considered as an abstraction of those facts of the case. The modular ANN is a candidate for this stage for its acclaimed ability to analyze complex problem and ability, though the capability of inference is limited. The second stage, legal elements identification, is equipped with C4.5 decision tree due to its inference logic which is needed for human-interpretation. The framework also considers the high dimensionality and limited availability of input data by preprocessors handling standardization, validation, and most importantly the factor analysis to dynamically and context-sensitively decrease the input dimensions subject to the power of each input data in explaining the output variance.

Now, the framework is in the process of transformation into a complete identification system for specific part of the Thai's criminal law codes. Once finished, the performance of the system shall be measured in terms of model accuracy, contingency analysis, and interpretability. The measures shall provide other insights into the domain of methodology and law concerns, e.g. the model adaptability to the limitation of data, the sensitivity of the sentence to some major factors, and the interaction of contingency to the model accuracy (Receiver Operating Characteristics, ROC) [21].

References

[1] J. G. Apple and R. P. Deyling, A Primer on the Civil-Law System, Federal Judicial Center, Washington D.C., 1995.
[2] N. B. Bilgi and R. V. Kulkarni, An Investigative Survey of Application of Knowledge Based System in Legal Domain, Int. Journal of Information Technology and Knowledge Management ,2008, pp. 517-525, ISSN 0973-4414.
[3] K. D. Ashley, Case-Based Models of Legal Reasoning in a Civil Law Context, International Congress of Comparative Cultures and Legal Systems of the Instituto de Investigaciones Juridicas, 2004, Universidad Nacional Autonoma de Mexico, Mexico City.
[4] H. Deng, G. Runger and E. Tuv, Bias of importance measures for multi-valued attributes and solutions, Proc. of the 21st Int. Conf. on Artificial Neural Networks, 2011, pp. 293–300.
[5] J. Gutierez and G. Leroy, Using Decision Trees to Predict Crime Reporting, Advanced Principles for Improving Database Design, Systems Modeling, and Software Development, IGI Global, 2008, pp. 132 - 145.
[6] R. Yang, S. Olafsson, Classification for Predicting Offender Affiliation with Murder Victims, Expert Systems with Applications, 2011, Vol.38, No.11, pp. 13518-13526.
[7] A. Stranieri and J. Zeleznikow, Knowledge Discovery from Legal Databases, Law and Philosophy Library, 2005, Vol. 69, Springer.
[8] J. Zeleznikow, Using Web-based Legal Decision Support Systems to Improve Access to Justice, Information & Communications Technology Law, 2002, Vol. 11, No. 1, pp. 15-33.
[9] G. C. Oatley, B. W. Ewart and J. Zeleznikow, Decision Support Systems for Police: Lessons from the Application of Data Mining Techniques to 'soft' Forensic Evidence, Proc. The 24th SGAI Int. Conf. on Knowledge Based Systems and Applications of Artificial Intelligence, 2006, pp. 35-100.

[10] M. M. J. Theresa and V. J. Raj, Analogy Making in Criminal Law with Neural Network, Proc. Int. Conf. Emerging Trends in Electrical and Computer Technology, 2011, pp.772-775.
[11] S. C. Watanapa, B. Thipakorn, and N. Charoenkitkarn, A Sieving ANN for Emotion-Based Movie Clip Classification. Trans. in Information System, 2008, Vol. 5, pp. 1562-1572.
[12] J. Han and M. Kamber, Data Mining: Concepts and Techniques, 2nd edition, Morgan Kaufmann, 2006.
[13] G. H. Dunteman, Principal Component Analysis, Springer, Sage Publications, 1989.
[14] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene Selection for Cancer Classification using Support Vector Machines. Machine Learning, 2002, pp. 389-422.
[15] P. A. Devijver, and J. Kittler, Pattern Recognition: A Statistical Approach, Prentice-Hall, London, 1982.
[16] M. Craven and J. Shavlik, Using Neural Networks for Data Mining, Future Generation Computer Systems 13, 1997, pp. 211-229.
[17] A. J. Sharkey, Combining Artificial Neural Nets: Ensemble and Modular Multi-Net Systems, 1st edition, Springer-Verlag New York, Inc., Secaucus, NJ, USA.
[18] J. R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, 1993.
[19] J. R. Quinlan, Induction of Decision Trees. Machine Learning, 1986, Vol. 1, No. 1, pp. 81-106.
[20] Y. V's Yuthankun (translated), The Criminal Code: Translated Thai-English, Soutpaisal Publisher, Bangkok, 2009.
[21] T. Fawcett, An introduction to ROC analysis, Pattern Recognition Letters, 2006, Vol. 27, No. 8, pp. 861-874.